APPLYING HYPERSPECTRAL IMAGING TO HEART RATE ESTIMATION FOR ADAPTIVE AUTOMATION

THESIS

Joshua M. Splawn, Captain, USAF

AFIT-ENV-13-M-30

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENV-13-M-30

APPLYING HYPERSPECTRAL IMAGING TO HEART RATE ESTIMATION FOR ADAPTIVE AUTOMATION

THESIS

Presented to the Faculty

Department of Systems and Engineering Management

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Systems Engineering

Joshua M. Splawn, BS

Captain, USAF

March 2013

AFIT-ENV-13-M-30

APPLYING HYPERSPECTRAL IMAGING TO HEART RATE ESTIMATION FOR ADAPTIVE AUTOMATION

Joshua M. Splawn, BS
Captain, USAF

Approved:

Michael E. Miller, PhD (Chairman)	Date
Jeffrey D. Clark, Lt Col, USAF (Member)	Date
Brent T. Langhals, Lt Col, USAF (Member)	Date

Abstract

The use of automation continues to increase in Air Force (AF) systems with the goal of improving operator efficiency and effectiveness. Unfortunately, systems are often complex, potentially imposing increased mental task load on the operator, or placing the operator in a supervisory role where they can become overly dependent on automation. A proposed solution is adaptive automation, where automation is triggered when an operator is overloaded, and disabled as the operator is underloaded. Changes in physiological measures have shown promise in triggering changes in automation levels. Unfortunately, physiological measurement techniques, such as heart rate measurement, can be obtrusive and impractical in day-to-day operations. This research used the Air Force Multi-Attribute Task Battery (AF_MATB) to impose varying task loads on thirteen subjects while monitoring their performance, recording their heart rate information with an electrocardiogram (ECG) and obtaining subjective estimates of mental workload. Simultaneously, hyperspectral images were captured to determine if changes in heart rate might be identified through these images, providing a remote assessment of heart rate. Heart rate (HR) and several heart rate variability (HRV) measurements where significantly affected by Task Load. A linear regression model was developed to predict subjects' perceived workload as a function of a proposed summary performance metric and HR measures. Additionally, this research identified several requirements for future remote HR monitoring techniques.

I would like to dedicate this thesis to my beautiful wife. Thank you for always supporting me! This whole program would have been much less enjoyable without you by my side.

Acknowledgments

First of all I would like to thank my advisor, Dr. Michael Miller, for all of his support and guidance throughout this process. I appreciate your patience and feedback throughout each phase of the research.

I would also like to thank the folks at the 711th HPW/RHCPA for providing the research funding, as well as guidance in regards to analyzing the heart rate measures.

Joshua M. Splawn

Table of Contents

Page

List of Figures

List of Tables

APPLYING HYPERSPECTRAL IMAGING TO HEART RATE ESTIMATION FOR ADAPTIVE AUTOMATION

I. Introduction

General Issue

Air Force (AF) systems, in general, have become increasingly complex and require the operator to perform more tasks than in times past. In particular, Remotely Piloted Aircraft (RPA) have seen explosive growth and are one of the most demanded capabilities that the USAF presents to the Joint Force (USAF, 2009). Along with the surge in missions, battlefield operations are becoming increasingly complex and automated as new technologies are rapidly introduced to combat changing threats. Often, operators can still function at an optimum level; however, if the task load becomes too great operator performance can suffer (de Waard, 1996). Consequently, technologies and techniques are needed to assist RPA operators in maintaining optimal performance.

One possible method of aiding operators is through adaptive automation (AA). AA is defined as "a system of controlling flexibly and dynamically the allocation of tasks between human operators and computer systems in complex multi-task environments" (Tattersall & Fairclough, 2003). Parasuraman discusses five methods to trigger this automation, which include: critical events, operator performance, operator physiological assessment, operator modeling, and a hybrid of these techniques (2003). The two techniques that have shown promise in triggering changes in adaptive automation within several studies are operator performance and physiological measures (Parasuraman, 2003). However, these techniques require continuous monitoring of both operator performance and physiological state to determine the operator's mental workoad level.

Throughout the years, human factors engineers have employed many techniques to quantify and measure workload. These techniques vary greatly, from purely subjective questionnaires (of which there are many), that require the participant to measure their own perceived workload, to objective physiological measures such as cardiovascular responses, oculometry, galvanic skin response, and even fMRIs (Booher, 2003). In particular, cardiovascular measurements, such as heart rate (HR) and heart rate variability (HRV), have shown correlation with mental workload levels in many studies (Averty, Athenes, Collet, & Dittmar, 2002; de Waard, 1996; Murai & Hayashi, 2011; Verwey & Veltman, 1996; Wickens & Hollands, 2000). However, standard methods of obtaining heart rate measures are intrusive and are not suitable for day-to-day RPA operations.

In recent years technologies have been developed that use various imaging techniques to monitor heart rate. The traditional imaging technique, known as photoplethysmography (PPG), uses light reflectance from the skin and has been used to monitor oxygen saturation (pulse oxymetry), heart rate (HR), and respiration rates (RR) (Verkruysse, Svaasand, & Nelson, 2008). Although this technique does not monitor electrical activity, as is typical for traditional cardiovascular measures, it still requires the device to have contact with the operator. However, several researchers have developed methods to remotely measure these signals with commercial imaging devices (Verkruysse, Svaasand, & Nelson, 2008; Kamshilin et al, 2011), spectrometers (Corral, Paez, & Strojnik, 2012), and hyperspectral imagers (Yuen, et al., 2009) within controlled environments and with controlled operators, which opens the door to remotely and unobtrusively measuring an operator's heart rate.

Imaging techniques have the potential to remotely measure an operator's heart rate, which has shown correlation with mental workload levels, in daily operations, allowing adaptive automation to handle some of the workload. The benefits of using PPG to measure heart rate are very promising, but these techniques do come with their own set of limitations that must be assessed to determine their suitability in various environments.

Problem Statement

Specific heart-rate measures that correlate to mental workload need to be identified across a range of individuals to provide data for future adaptive automation systems. The development of remote-monitoring techniques (e.g. hyperspectral imaging and PPG) provides multiple options to implement continuous physiological monitoring in everyday situations. However, each technique has its own limitations that need to be identified and analyzed to determine their usefulness in various applications.

Research Objectives

The two primary objectives of this research are to: 1) Provide insight into heart-rate measures that can be applied to adaptive automation systems and 2) Identify, analyze, and compare the different monitoring techniques that are available and provide recommendations on their suitability. A secondary objective is to partner with researchers in the Electrical and Computer Engineering Department to further develop a method of capturing heart rate with hyperspectral imaging techniques.

Investigative Questions

This thesis will strive to provide answers to the following questions:

1) What are the heart rate measures that indicate an operator has reached task saturation?
2) What are the variables affecting the robustness of heart rate measurement techniques and what are some of the requirements to implement these techniques in an adaptive system?
3) Can we develop a robust hyperspectral technique to remotely detect heart rate?

Methodology

To adequately investigate the first and third research questions, this study will perform human subjects testing. The testing will utilize the Air Force multi-attribute task battery (AF_MATB) software program to impose varying task loads on participants, while heart rate data is recorded with an electrocardiogram (ECG). Task loads will vary from low workload through high workload levels with two moderate levels of workload in between. Performance will also be measured and compared to HR and HRV to determine what heart rate measures are indicative of decreased performance. Hyperspectral data collects will be accomplished throughout the testing to search for a correlation between the hyperspectral data and the heart rate measurements.

The second research question will be investigated through a review of relevant literature. The review will be conducted to identify the variables that affect and are affected by the different heart rate measurement techniques. Requirements for successful

implementation will also be identified in the literature and investigated through the results of the research.

Assumptions/Limitations

The research will be limited by the number of individuals who will be tested and can be recruited. There is an assumption that a trained operator's heart rate will vary similarly to those that are tested for this research. The research is also limited by the equipment that is available. Another limitation is the time available for testing, not only in terms of data collection period, but also the time that each subject can contribute as they will be volunteers who also have busy schedules.

Implications

The development of a robust remote heart-rate measurement technique will give the AF a better understanding of operator stress levels, workload levels, and overall health in applications where it has previously not been implemented due to its obtrusiveness. Implementing the findings of this research with further AA development will provide resources to aid in maintaining current RPA operations without negatively affecting performance.

II. Literature Review

Chapter Overview

This chapter seeks to bring together research on three focus areas: adaptive automation, mental workload, and remote heart rate measurement. The high-level goal is

to successfully implement adaptive automation to increase overall mission performance in complex systems. An important aspect of invoking this automation is determining when an operator is struggling to perform all required tasks and/or they are experiencing high mental workload. Past experiments have shown that heart rate (HR) and heart rate variability (HRV) are indicative of high workload levels. Studies have also used changes in HR and HRV as triggers to invoke automation and have shown success in increasing operator performance compared to a control group. The last consideration is the impracticality of having to use an ECG to record heart rates in day-to-day operations. One possibility to increase operator acceptance is through remotely monitoring their heart rates. Remote measurement has been successfully implemented using various PPG methods in controlled environments, but there are other types of imaging, such as hyperspectral imaging (HSI), that may provide more robust measurement techniques.

Relevant Research

Adaptive Automation

Modern systems have become increasingly complex and allow more tasks to be accomplished while reducing the number of personnel needed to operate these systems. However, to successfully operate these systems, many of the sub-system tasks have been automated, placing the user in a supervisory role (Parasuraman , 2003). The goal of the automation is to increase efficiency while lowering costs and decreasing the task loads placed on the human operator. While this is often achieved there can also be costs in terms of human and/or system performance (Parasuraman, 2003). These costs can include reduced situation awareness, decreased system performance due to lack of trust in

automation, cognitive skill degradation, and unbalanced mental workload (Parasuraman & Byrne, 2003). Therefore, the goal is to develop a system that retains the benefits of automation while minimizing the costs incurred when automation is implemented poorly. One way to accomplish this is through a system in which the automation and the operator work together to accomplish the overall objective.

"Adaptive automation represents an approach to automation in which the allocation of functions to human and machine agents is not set inflexibly at the design stage but is changeable during system operation (Parasuraman & Hancock, 2008)." This concept was first discussed in 1976 and the idea has continued to evolve throughout the years. However, the technology to effectively implement this solution has only recently begun to mature (Parasuraman & Hancock, 2008). A key concern with adaptive automation is the classic problem of properly allocating functions to the human and the system. Because allocations cannot always be accurately made based on stereotypical characteristics of human and computer capabilities, either the system or the operator must be tasked with guiding this allocation during system operation. Requiring the operator to guide this operation, however, adds to the workload of the user, which can be problematic at the times when increased automation is required to reduce human workload. Therefore, there is a need for an adaptive computer aid that responds to task demands and operator performance (Parasuraman & Byrne, 2003). The design of this adaptive computer aid is affected by the method by which adaptation is implemented in the system.

A number of methods for adaptive automation have been proposed; however, (Parasureaman and Byrne 2003) conclude that the major techniques fall into five main categories:

1. Critical events
2. Operator performance measurement
3. Operator physiological assessment
4. Operator modeling
5. Hybrid methods

In the critical-events method, automation is implemented based on the occurrence of specific tactical events that would normally impose an increased task load on the operator (Parasureaman and Byrne, 2003). This method can therefore be flexible based on current tactics and doctrine for mission planning, and it is relatively easy to implement (Parasuraman & Hancock, 2008). However, the critical-event method may be insentive to operator and or system performance and may become a nuisance, as this method will invoke automation whether or not the operator requires aid when the critical event occurs (Parasuraman & Hancock, 2008). Another disadvantage of this technique is that it may not be possible to accurately identify all critical events during the planning phase, resulting in a system that does not invoke automation to reduce operator workload during these critical events.

The operator performance measurement and physiological assessment are often discussed as one integrated method in the literature because they both deal with measuring the operator’s state. The goal is to accurately determine when the operator is experiencing high mental workload by measuring their performance and different

physiological measures, such as Electroencephalography (EEG), Electrocardiography (ECG), Electrooculography (EOG), and respiration that correlate with workload, and to then invoke automation when the operator's task load is too great. This technique has the advantage of enabling the system to automatically adapt to the current operator's needs depending on the situation. Unfortunately, the use of these measures is reactive so the operator may already be overloaded before these measures indicate a change in user performance or physiological state. In the literature, these two methods appear to have shown the most success as they are responsive to the needs of the operator.

In a 2003 study, Parasuraman conducted a study in which two groups of participants used a modified version of MAT-B, called EICAS-MAT, for 90-minute sessions. The session was divided into three consecutive phases of high, low, and high difficulty with adaptive automation being provided for the high difficulty phases for one group; however, this automation was only triggered when heart rate variability (HRV) was reduced below a specific point (Parasuraman, 2003). The second group did not receive any aid from the automation regardless of their performance or heart rate measures. The results showed that the adaptive group had generally higher HRV, which is indicative of lower workload, and that their tracking performance was superior to the control group (Parasuraman, 2003). However, the operators' performance on the remaining three three tasks within this simulated envrionment were not discussed. Although the implications of this omission is uncertain, one might assume that their was no difference between groups in those tasks.

Another study by Parasuraman, Cosenzo, and De Visser conducted in 2009 researched the effect of adaptive automation that was triggered based on operator

performance of one of the required subtasks. The primary task was controlling an unmanned ground vehicle (UGV) that was in contact with an UAV. Participants were required to complete specific actions at various waypoints depending on the situation (Parasuraman, Cosenzo, & De Visser, 2009) as well as complete two communications tasks. One of the communication tasks required operators to acknowledge their call sign if presented and ignore others, while the second communications task was a situation awareness probe that asked varying yes/no questions. The final task was a change detection task during which various icons would unpredictably change locations throughout the scenario and the participants were required to hit the space bar when they noticed a change (Parasuraman, Cosenzo, & De Visser, 2009). The operator's performance for this task was used to determine whether or not automation should be invoked. The results showed that the adaptive automation increased performance on only the fourth task and it also reduced the subjective assessment of workload. However, there were no significant differences for the other tasks.

The final two methods of invoking automation are operator modeling and a hybrid of the techniques. A method to avoid high workload conditions is through modeling the possible future workload states of the operator to determine when automation should be invoked (Parasuraman & Hancock, 2008). While there have been some initial studies that have used the Improved Performance Research Integration Tool (IMPRINT) to model an operator's workload, the tool has not been thoroughly validated and the model would need to be adjusted for each operator and scenario. One way to minimize the impacts of the disadvantages in each of these techniques is through an integration, or "hybrid", of the various methods that seeks to plan for the critical events through

modeling, while also reacting to the current situation and operator with physiological and performance measurement (Parasuraman & Hancock, 2008).

Mental Workload

To effectively implement adaptive automation, it is important to understand mental workload and methods to measure it. There are several definitions of workload in the literature, but the underlying theme is that mental workload is the amount of mental effort needed to accomplish a task or goal. One definition determines workload to be "the amount of cognitive or attentional resources being expended at a given point in time" (Booher, 2003), while another states, "Workload is a general term used to describe the cost of accomplishing task requirements for the human element of man-machine systems (Tsang & Vidulich, 2003)." While these two former definitions are relatively similar and focus on the resources being expended by the operator of the system, another definition looks at workload as the task demands placed on the operator (Wickens & Hollands, 2000).

There are a number of criteria that should be considered when implementing a workload assessment technique: 1) sensitivity to changes in task difficulty or resource demand; 2) diagnosticity to determine between which mental resources are being taxed; 3) obtrusiveness to primary task performance (Wickens & Hollands, 2000). Booher (2003) discusses the four traditional methods of measuring workload, which will be described with a focus on the techniques which can be more objectively measured (1,2, and 4) as these are the most applicable for use in adaptive automation. The four techniques are:

1. Performance on primary task measures;

2. Performance on secondary task measures;

3. Subjective measures; and

4. Electrophysiological and psychophysiological measures.

Primary Task Measures

When using primary task measures an operator is given tasks to complete and specific emphasis is placed on the most important task. According to Booher speed and accuracy is stressed for that test and operator workload level is derived from performance on that primary task (Booher, 2003). This technique is useful because it can be applied either in a laboratory controlled experiment or in an operationally representative setting where post-flight data can be analyzed. One study found that this technique is sensitive to different durations (15s, 30s, and 60s) of workload (Verwey & Veltman, 1996). However, there may be other factors influencing performance, such as poor design layout, lack of feedback, inaccurate mapping between controls and displays, or simply the user having a poor mental model of the system. According to Wickens, this technique is also not sensitive to “underload” conditions where the operator’s reserve mental capacity has not been depleted and they can still successfully accomplish the task (Wickens, 2000). This relationship amongst resource supply, resource demand, and primary task performance can be seen in Figure 1. As shown, primary task performance can be expected to be high as long as the resources demanded are less than the maximum available resources, leaving excess resource capacity. However, once excess resource capacity is depleted primary task performance decreases.

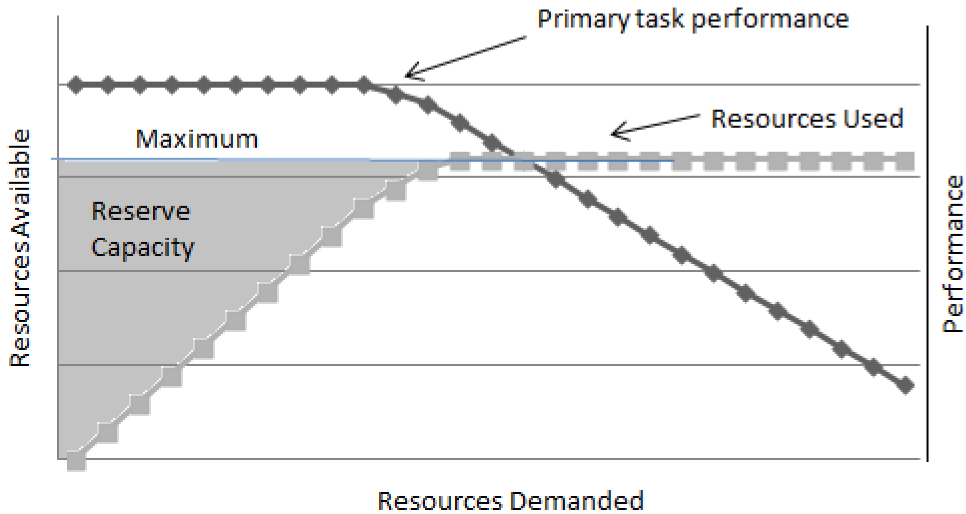

Figure 1: Task performance vs resources demanded (Adapted from Wickens, 2008)

Secondary Task Measures

The second workload measurement technique is to measure performance on secondary task measures. The logic behind this technique is that operators have limited processing resources, and they will only be able to allocate resources to the secondary task when they have reserve capacity (Tsang & Vidulich, 2003). Therefore, when primary task demands are low, secondary task performance will be high, and when primary task demands are high the secondary task performance will be relatively poor (Tsang & Vidulich, 2003). This appears to be a simple, straightforward measurement technique and a study by de Waard showed that it is sensitive to changes in workload levels (de Waard, 1996). However, there are some difficulties that may arise when employing secondary task measures to evaluate workload. The first difficulty relates to Multiple Resource Theory (MRT), which asserts that operators can employ multiple

mental resources simultaneously (Wickens, 2008), so it is essential that the secondary task uses the same mental resources as the primary task.

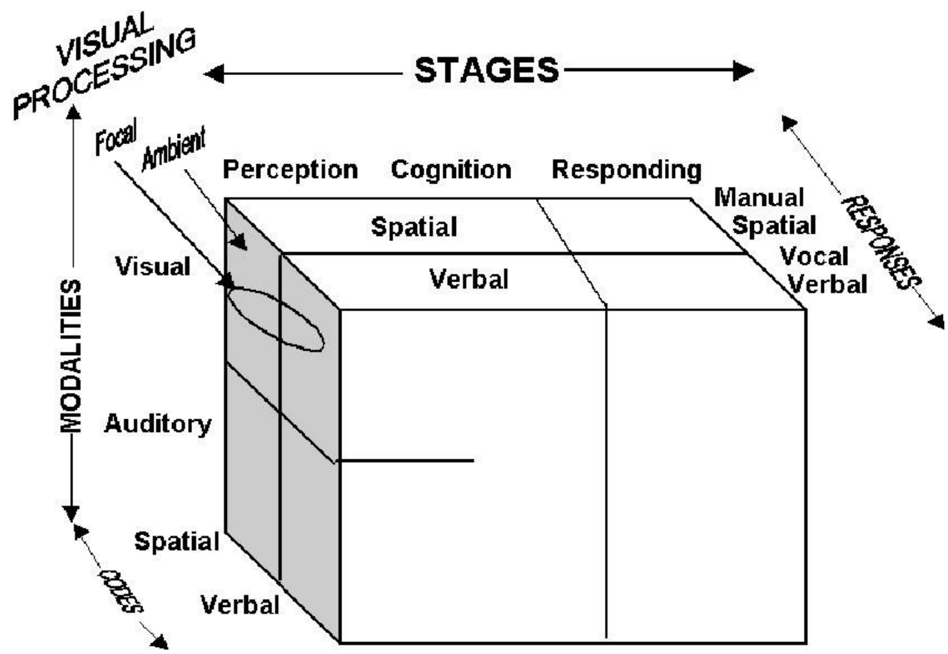

Figure 2: Model of Multiple Resources Theory (Wickens, 2008)

Vidulich states that another potential problem with employing this technique is that the secondary task can often have a negative impact on primary task performance (Vidulich, 2003). This assertion was confirmed in a study by Verwey and Veltman that compared nine different workload assessment techniques. In this study, it was found that even though the secondary task was sensitive to workload, it also negatively affected primary task performance (Verwey and Veltman, 1996). Another difficulty in applying this technique, especially in an operational system, is that if the primary task imposes a high level of workload or if the operator is already required to multitask, the secondary task may not be performed at all and will not be able to discriminate between different levels of high workload.

Subjective Measures

"Self-report measures have always been very appealing to many researchers. No one is able to provide a more accurate judgment with respect to experienced mental load than the person concerned" (de Waard, 2006). This method of collecting subjective ratings from operators has high face validity, is intuitive, and is also easy to use (Booher, 2003). There are also many questionnaires to choose from (Stanton, 2005), and many studies have shown that they can be both sensitive to workload levels and diagnostic with respect to type of workload (Averty, Athenes, Collet, & Dittmar, 2002; Reid & Colle, 1988; de Waard, 1996; Verwey & Veltman, 1996). Another benefit of this technique is that it does not disrupt primary-task performance because it is completed post task (Wickens & Hollands, 2000). However, there are also limitations to this technique. Often, the operator may not clearly remember the specific task. Operators may also unconsciously employ one or more workload mitigation techniques (Delay, Avoid, Transfer, Simplify), which can lead to rating the workload level as lower that it actually was. Stanton et al also found that subjective ratings of workload are often correlated with performance level (Stanton et al., 2005). For example, if the operator performed well, they often rated the workload as low and vice-versa.

Physiological Measures

The final technique employed to assess mental workload is to record, unobtrusively, the effects of workload through appropriately chosen physiological measures (Wickens & Hollands, 2000). This technique is appealing, because it does not negatively impact performance, data is gathered continuously, and the measures can usually be correlated with workload. There have been many physiological metrics that

can be a good indicator of workload level; however, this research will primarily focus on cardiovascular measures. A specific cardiovascular measure that has been used in multiple studies is heart-rate variability (HRV). "In general HRV decrease is more sensitive to increases in workload than heart rate (HR) increase (de Waard, 1996)." Many studies have found that HRV is sensitive to a number of different difficulty manipulations (Averty, Athenes, Collet, & Dittmar, 2002; de Waard, 1996; Murai & Hayashi, 2011; Verwey & Veltman, 1996; Wickens & Hollands, 2000).

In the article, *Detecting Short Periods of Elevated Workload: A Comparison of Nine Workload Assessment Techniques*, Willem B. Verwey and Hans A. Veltman compared nine different workload techniques to determine their usefulness in measuring workload in a car-driving task. The authors placed special emphasis on both the sensitivity and diagnosticity of the workload assessment techniques. Sensitivity is whether the technique discriminates between levels of workload and diagnosticity is whether the technique distinguishes between types of workload (i.e. visual workload or mental workload) (Verwey & Veltman, 1996).

The goal of the study was to compare both the sensitivity and the diagnosticity of a series of workload measures for short (10s, 30s, and 60s) periods of elevated visual and mental workload (Verwey & Veltman, 1996). The method used to accomplish this, had 12 experienced, male drivers operate an instrumented car on a 40-km, four-lane freeway that is characterized by stable traffic with little or no congestion that took approximately 30 minutes to drive (1996). The task involved lane keeping and controlling speed and to increase workload the experimenters added a visually loaded task and a mentally loaded task to the driving task. These loading tasks lasted for 10, 30, or 60 seconds. These

longer periods of elevated workload were introduced to obtain an indication of the sensitivity and diagnosticity of the workload assessment techniques (Verwey & Veltman, 1996).

Nine workload assessment techniques were then used to assess the workload of the drivers. Two involved primary task measures that are known to be affected by driver workload (Verwey & Veltman, 1996). The experimenters also employed two subjective workload assessment techniques, which were the Subjective Workload Assessment Technique (SWAT) and the Rating Scale Mental Effort (RSME). Finally, the authors employed four different physiological measures for workload: interbeat interval (IBI), HRV, intervals between successive eye blinks, and intervals between successive skin conductance responses (SCRs) (Verwey & Veltman, 1996).

With regard to physiological measures, IBI did not provide any significant indication as to the workload level. However, HRV, which was computed by dividing the standard deviation of the IBIs by the average IBI, was a significant indicator for the 60-second duration tasks and even showed a difference between the no-driving and control-driving tasks.

In his 1996 dissertation, de Waard recorded various performance, subjective, and physiological measures to determine which were sensitive to workload in driving tasks (de Waard, 1996). He looked at heart rate, IBIs and their variance, and the 0.1 Hz frequency for 22 subjects. He found that the frequency measure of 0.1 Hz was more sensitive than other heart rate measures to mental load manipulations.

A more recent study was performed that used MATB in the testing and looked at different heart rate measures as indicators of workload (Miyake, Yamada, Shoji, Takae,

Kuge, & Yamamura, 2009). In this study 15 participants performed three MATB trials in the same order and came back three days in a row to determine if heart rate measurements were consistent. The researchers looked at several HRV indices which included the low frequency (LF) component, the high frequency (HF) component, LF/HF, total power, and CV-RR, which is the standard deviation of IBIs divided by the average IBI (Miyake, Yamada, Shoji, Takae, Kuge, & Yamamura, 2009). They found that LF/HF did not correlate with difficulty level; however, they did find that the LF (0.1 Hz) component did show high test/retest correlations (Miyake, Yamada, Shoji, Takae, Kuge, & Yamamura, 2009).

Although this technique shows much promise, it still has a number of limitations. The primary limitation is that HRV is not only affected by mental workload, but also by physical workload (de Waard, 1996; Novak, Mihelj, & Munih, 2010). Another limitation is operator acceptance of the measures, as it would be uncomfortable to be hooked up to an ECG for day-to-day missions.

Remote Heart Rate Measurement

One of the previously mentioned limitations of utilizing physiological measurements to control adaptive automation is that these measures can become intrusive and, therefore, impractical for standard operations. While this may continue to be a limitation for EEG and fMRI measurements, there has been research into remote measurement of an individual's heart rate through various imaging techniques.

The oldest of these imaging techniques is photoplethysmographic (PPG) imaging, which has been used to detect the blood pulse travelling throughout the body. This method utilizes a light source and examines the changes in reflected energy due to blood

flow. Researchers have discovered that a correlation exists between the intensity modulation of reflected light from the skin and a person's heartbeat (Kamshilin, 2011). Initial studies allowed researchers to visualize dynamic changes in cardiovascular pulse wave, resulting in a distinguished heartbeat (Kamshilin, 2011).

Other studies have been able to remotely sense heart and respiration rates without a dedicated light source (Verkruysse, Svaasand, & Nelson, 2008). This method used images from the forehead and can extract the signals as the intensity of reflected light changes; however, this technique is highly sensitive to movement as it changes the area being measured (2008).

Hyperspectral imaging (HSI) is another imaging technique that allows for more discriminating features to be examined, which may allow for a more robust detection method. Pan and colleagues have demonstrated that with HSI, deeper skin layers can be imaged, producing more distinguishing results than may occur at surface level collection (Pan et al., 2003). As blood pulses through the body, the reflectance of the skin changes, providing the possibility of detection using HSI by implementing techniques similar to those of PPG which uses the ratio of select narrowband reflectance values to determine blood oxygen content.

Summary

Based upon the background literature, it is believed robust method of remote heart rate measurement would be an excellent tool to implement adaptive automation in more operational settings. It would allow for real-time operator monitoring, which could be used to trigger automation when heart rate measures indicate that the operator is

experiencing high workload. This measure might be combined with other measures, including operator performance to provide a robust trigger for automation changes, especially if a method could be provided to permit the remote measurement of heart rate variability and associated measures.

III. Methodology

Chapter Overview

To adequately investigate the first and third research questions, this study used human subjects testing. The testing utilized the Air Force multi-attribute task battery (AF_MATB) software program to impose varying task loads on participants, while heart rate data was recorded with an ECG. Task loads varied from low through high with two moderate levels of task load. The subject's perceived mental workload and performance were measured and compared to heart rate and heart rate variability to determine HR measures which are indicative of decreased performance. Hyperspectral data collects were accomplished throughout the testing to search for a correlation between the hyperspectral data and the heart rate measurements.

Equipment

Subjects interacted with the Air Force Multi-Attribute Test Battery (AF_MATB), running on a laptop computer. The AF_MATB provided a method to manipulate an operator's task load and impose different levels (high, med, low) mental workload (Miller, 2010). The original MATB software has become a mainstay for psychological research regarding cognitive workload and this version has simply updated the software

to be compatible with modern operating systems (Miller, 2010). Subjects used the standard laptop keyboard in addition to a USB joystick to perform the given tasks.

Integrated within AF_MATB is a subjective workload assessment scale that was used to measure the subject's workload. The scale the subjects used is the NASA Task Load Index (TLX), which has been used in many studies as an effective means to measure subjective impression of workload (Miyake, Yamada, Shoji, Takae, Kuge, & Yamamura, 2009; Stanton, 2005; Kawakita, Itoh, & Oguri, 2010). The six components of the NASA TLX scale used to provide an aggregate measurement of workload are mental demand, physical demand, temporal demand, performance, effort, and frustration. A copy of this index is provided in Appendix A.

An Analytical Spectral Devices (ASD) FieldSpec® Pro spectrometer was employed in the experiment to measure spectral bands that are associated with heart rate. The FieldSpec3 measures wavelength, absolute reflectance, radiance, and irradiance. The spectral response was collected with a contact probe that rests against the subject's skin and a remote device.

Subjects wore electrodes on their chest as indicated in Figure 3. These electrodes were attached to a BIOPAC 150 with an ECG 100C amplifier (See Figure 4) for measuring the electrical signals associated with the beat of the human heart. The output from this instrument was the truth source for heart rate.

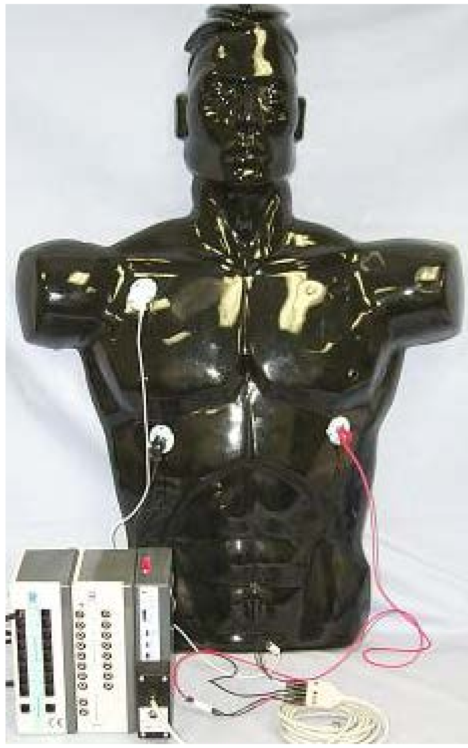

Figure 3: Electrode placement for ECG

Subjects

Thirteen subjects were recruited for this study. There were 4 females and 9 males. The average age of participants was 25 years old, and all were in good health. Ten participants had no prior experience with AF_MATB, while three had minimal prior experience. The number of participants was consistent (each study had twelve subjects) with previous studies that investigated the correlation between heart rate measures and workload levels (Veltman & Gaillard, 1996; Verwey & Veltman, 1996). The subjects that participated in this study were voluntary military and government civilian personnel. Subjects were not compensated for their participation.

Qualifications and Time Commitment

Testing consisted of one formal test session that lasted approximately one and a half hours. This included 10 minutes for setup, calibration, and study information, 15 minutes for software training and heart rate baseline, 35 minutes for the actual test runs, and 5 minutes for debrief.

Description of experiment, data collection, and analysis

Sensor Application

The subject was fit with the three ECG probes for at least 10 minutes prior to the actual testing to determine a baseline heart rate for the subject. To ensure that a good signal was received from the electrodes it was important to complete the following: 1) abrade the skin, apply electrode gel, 2) wait five minutes before collecting data to allow gel to soak in, and 3) minimize the participant's movement (BIOPAC System, 2000). The ECG fitting required the probes to be attached to the participant's chest and required clothing which covers the upper torso to be lifted or removed to permit probe placement. This fitting took place with the assistance of an individual having the same gender as the participant, while the other investigator left the room. Once the probes were fit, the participant was permitted to return their clothing to its proper position on their torso before the other investigator re-entered the room. After completion of the study, the subject was asked to remove the probes, clean any remaining contact gel and reposition any clothing.

Calibration

To ensure that a smooth and continuous ECG waveform was obtained from the hardware the ECG100C Amplifier (See Figure 4) had to be correctly set up.

Figure 4: BIOPAC MP150 and ECG100C Amplifier

According to guidelines from BIOPAC Systems Incorporated, the amplifier had to adhere to the following settings to ensure minimal waveform distortion: 1) Gain: 1000 samples per second, 2) Mode: NORM, 3) 35Hz LP: On, 4) HP: 0.5Hz (BIOPAC System, 2000). Once the subject was fit with the ECG probes a baseline heart rate was determined for the subject. For subjects 1-7, baselines were obtained when the

participants were listening to instructions without much control over their relaxation; however, the investigators realized the baselines were more variable than desirable. To ensure the next 6 participants were relaxed when determining the baseline measures they were shown a slideshow of calming images (Fedorovskaya, et al., 2001) and concentrated on controlled breathing while refraining from speaking. HSI was also accomplished during this baseline measurement.

AF_MATB Tasks

The AF_MATB consists of 4 types of tasks that simulate tasks analogous to those a flight crewmember would encounter. Subjects operate a joystick and keyboard to complete these tasks. Tasks include system monitoring, tracking, communication, and resource allocation (See Figure 5). System monitoring consists of monitoring four gauges and two lights, and the subject can provide corrective action via the keyboard when appropriate. The tracking task is performed with the joystick. The objective is to keep the unstable crosshairs within a designated rectangular target area. Communication requires the subject to listen for the appropriate call sign and then change the frequency for one of four channels via the keyboard. For the resource allocation task, subjects are responsible for turning on/off eight pumps to maintain a desired level (2500 +/- 300 for this research) in two main tanks in a constantly changing environment.

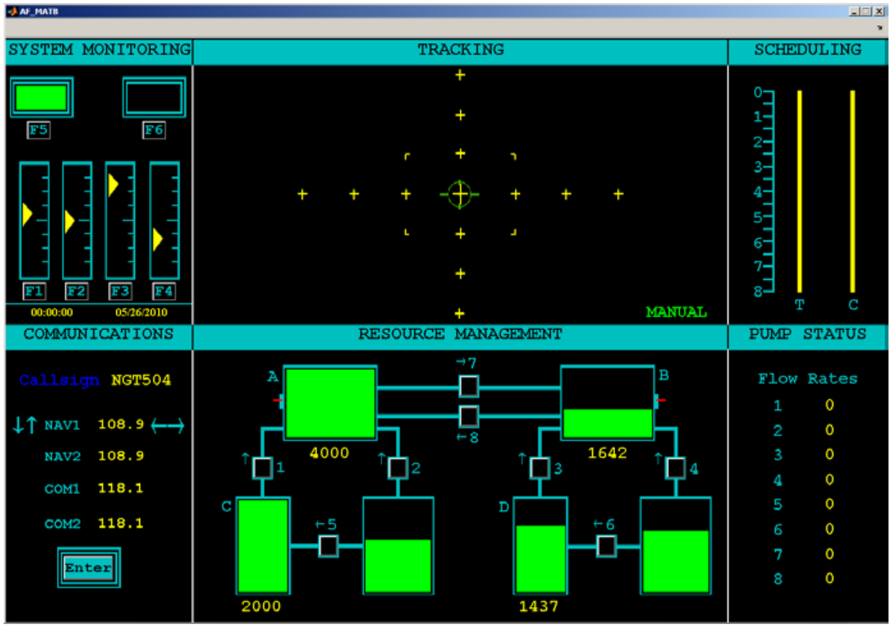

Figure 5: The operator interface for the AF_MATB tasks

Training

Subjects were shown the AF_MATB and given a brief overview of how it works and what they were required to do for each task. Subjects were then given up to 10 minutes to familiarize themselves with the software on various task load settings during which they were encouraged to ask questions about the operation of the software and the experimenter offered suggestions for improvement. At the end of the training, subjects were asked if they were ready to begin the experiment or if they wanted any extra time to practice.

Test Sessions and Testing Sequence

Prior to initiating the test session the subject was briefed on the study and its goals; were fit with the ECG contact probes; and underwent training as explained in the previous sections. The AF_MATB test scenarios began after baseline measurements were taken and the subject was trained. The subject accomplished three, five-minute

AF_MATB sessions at high, medium, and low difficulty levels. These difficulty levels were randomized to minimize a learning effect. Toward the end of each five-minute session a data capture was taken while the subject was still completing the task, and when the session ended the subject was asked to complete the NASA TLX subjective workload questionnaire. The subjects then repeated the process; however, this time the hyperspectral data was collected with the off-body sensor. Subjects were permitted to remove the ECG probes and were debriefed, which concluded their testing.

Data Collection and Analysis Preview

All performance data for the AF_MATB tasks was recorded and output by the AF_MATB software. This data includes the response times to errors for the system monitoring task and the communications task, but shows the root mean square (RMS) for the other two tasks; however, to measure overall performance the four individual task scores need to be rolled up into one global system score for each task load level. One performance score is desirable as it allows the investigators to look at correlations between the performance scores and the various heart-rate measurements. It also allows analysis of variance (ANOVA) to be applied to determine if the various task load levels significantly affect the performance scores. The AF_MATB also collects the subjective workload data input by the subjects and converts it to the NASA TLX weighted workload level (WWL), which is the subject's composite score for perceived mental workload. An ANOVA was performed on the WWL to determine if it is an effective method to accurately measure the operator's perceived mental workload.

The truth heart rate data, which includes heart rate and heart rate variability (HRV), was obtained from the AcqKnowledge software that is included with the

BIOPAC system. The software allows the data to be filtered, automatically locates the ECG waveforms, identifies their peak R waves, and then runs analysis to compute the different spectral components of HRV (LF, HF). The low frequency (LF) band also known as the '0.1 Hz component' is related to the short-term blood-pressure regulation and the high frequency (HF) band is believed to be influenced by respiratory-related fluctuations (de Waard, 1996). These two HRV components have been shown to be related to task demands and mental effort (de Waard, 1996). The AcqKnowledge software also outputs the inter-beat intervals (IBIs), also known as R-R Interval (RRI), which are the times between heart beats. The standard deviation of these RRIs, known as SDNN, divided by the average RRI value is known as CVRR (Coefficient of Variation of R-R) and has also shown correlation with task loads (Kawakita, Itoh, & Oguri, 2010). Two other factors that were investigated were HR delta and HRV delta. These measures were the amount the heart rate and low frequency component of HRV deviated from the baseline reading for each task load and were calculated in Excel. Each of these components of HRV along with standard heart rate was analyzed with ANOVAs to determine which measure(s) are most indicative of workload levels.

The spectrum output of wavelengths and reflectance values from the ASD were recorded and processed with RS^3 5.7 software. This software enables the user to optimize the FieldSpec3 instrument and collect various data types. This software converts the wavelength and reflectance data to a format that works with the computer program Matlab. Matlab was used to manipulate the data to distinguish the heart rate. The data from the ECG was also analyzed to determine whether significant changes in heart rate or heart rate variability occurred during each experimental session and these values were

correlated with changes in the reflectance data collected from the ASD. More information on this data collection and analysis is provided in a companion thesis (Norvell, 2013).

Summary

Human subjects testing was performed to gain insight into correlations between heart rate measures, performance, task load, and perceived mental workload. Thirteen subjects participated in the study with their individual sessions lasting approximately 1.5 hours. By analyzing the data and obtaining the results from this experiment more understanding of mental workload and heart rate measures should be obtained.

IV. Analysis, Results, and Discussion

Chapter Overview

ANOVAs were performed on operator performance, WWL, Heart rate, HR delta, HRV delta, the LF and HF components of HRV, SDNN, and CVRR to gain insight into which measures were affected by task loads and also to determine which measures were affected by the operator's experience with the AF_MATB tasks. Correlations between the measures were also investigated to determine which measures can be used to most accurately predict an operator's workload.

Data Processing

Certain metrics, such as heart rate and WWL were automatically calculated and output by the software (e.g. AcqKnowledge and AF_MATB) used in this study. However, the remainder of the data needed to be processed to some extent.

Heart Rate Data

In general, obtaining measures for both the time domain methods (e.g. SDNN, CVRR) and spectral methods (e.g. LF, HF) of HRV was relatively simple. The two most important considerations for HRV analysis were: 1) ensuring that the same time interval was measured for each subject and each task load, and 2) making sure the data was clean enough for the software to correctly interpret the ECG waveforms. The first consideration was simple to account for as the software allowed for selection of specific time intervals. A time interval of five minutes was chosen for this study based on guidelines from the American Heart Association (Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, 1996).

For the majority of the subjects (10 of the 13), the data had relatively little noise, which allowed the software to accurately identify the ECG waveforms. The data from the ECG electrodes was received and displayed by the AcqKnowledge software. Clean data appears as a typical ECG waveform as seen in Figure 6.

Figure 6: Clean ECG Signal

ECG can be susceptible to artifacts (See Figure 7) caused by excessive muscle movement.

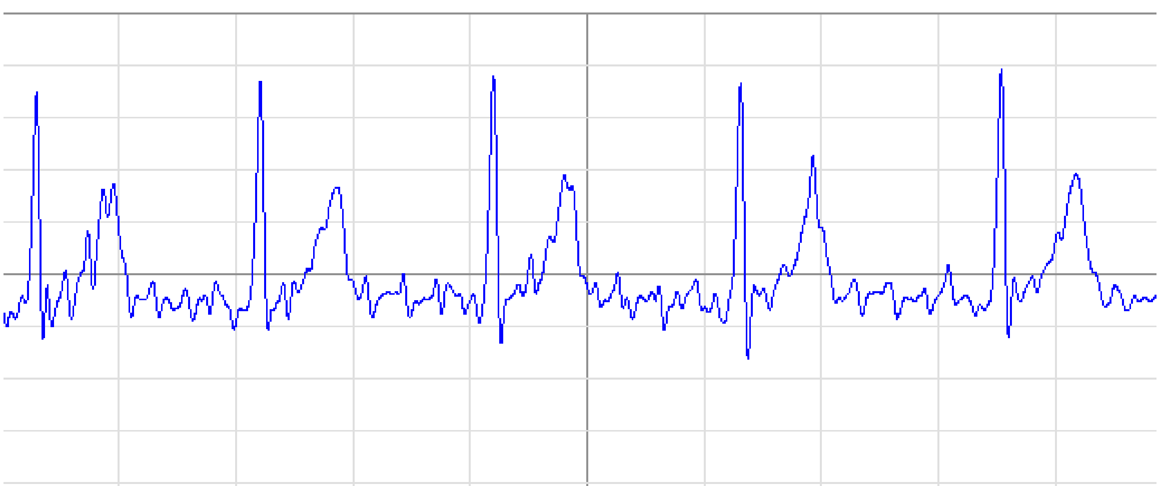

Figure 7: Noisy ECG Signal

Sometimes the software will mistake these signals for waveforms and other times the interference signals would disrupt part of a waveform, which would cause the

software to not count the waveform as a heartbeat. "Results reveal that even a single heart period artifact, occurring within a 2-min recording epoch, can lead to errors of estimate heart period variability that are considerably larger than typical effect sizes in psychophysiological studies (Berntson & Stowell, 1998)." By looking at a tachogram (See figure 8), which is a plot of the IBIs, it is relatively simple to detect anomalies. These graphs should be relatively sinusoidal without steep increases or decreases between each beat. Rapid changes indicate errors (either missed beats or additional beats) in the ECG data.

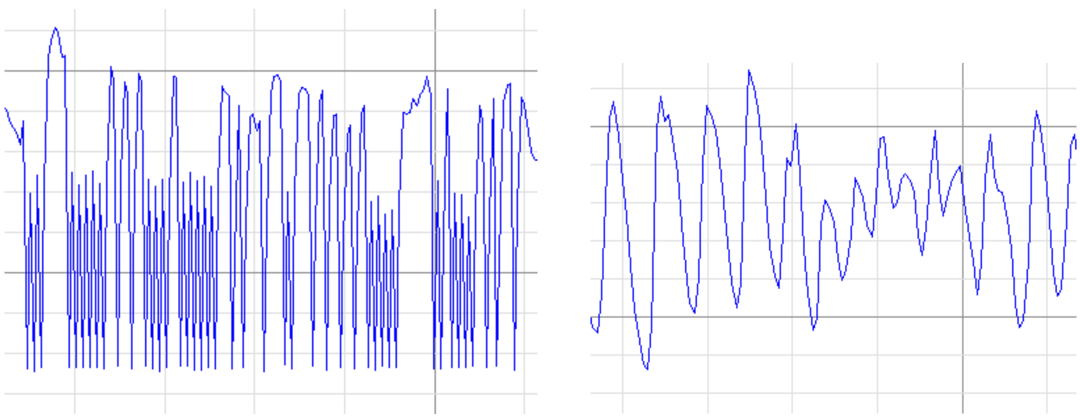

Figure 8: Noisy Tachogram vs Clean Tachogram

The AcqKnowledge software then allows for manipulation of the waveforms to cancel out noise and provides an alternative method of performing the HRV analysis (i.e. instead of detecting the entire waveform the software will compute the IBIs by counting the waveform peaks). By completing these data processing steps it was possible to obtain reliable HRV data.

Performance Data

One of the more difficult aspects of data processing was developing a global performance measured that could be easily compared between subjects. The literature investigated only looked at the performance for each of the four individual AF_MATB tasks separately, and while that approach was considered for application in the present study, it did not allow overall performance to be related directly to perceived workload or to the heart rate measures. A previous MAT-B analysis method used baud rates to quantify each task in similar terms; however, the research only developed this method for 3 of the four (Phillips, Repperger, Kinsler, Bharwani, & Kender, 2007). Because this current research was looking for correlations between performance, task load, and heart-rate measures a method to obtain an overall system score was developed.

The initial method developed for this research determined the percentage of time each subtask was in an incorrect state in relation to the entire task time. However, this method did not take into account the fact that as the workload levels increased the users had to respond to increasing numbers of system-induced errors. This increase of errors with difficulty levels, along with the fact that the system resets faults after 15 seconds if there is no operator input, made it impossible for the performance score to increase or remain constant as difficulty levels increased, despite the participants' performance. However, it intuitively makes sense that an operator could perform better at slightly higher difficulty levels if they were bored during really easy tasks and casually responded to system errors. To account for that possibility, each task was looked at as a percentage of time the task is in a correct state versus the total time it could be in a correct state. The total system score was then computed as the average percentage across the four tasks.

The computation was slightly different for the four tasks, because system monitoring and communications tasks are measured in terms of response times within the software while the root mean square (RMS) is recorded by the software for the tracking and resource management tasks and the time the RMS value was outside of allowable limits had to be determined from these RMS values.

For both the system monitoring task and the communication tasks the formula that was used to first determine the percentage of incorrect system time is: $X_i = \frac{(R_t * C)+(N_t * t)}{(E * t)}$. Where: X is the percentage of time the system is in an incorrect state, R_t is the average response time, C is the number of correct responses, N_t is the number of times the system timed out without an user response, t is time it takes the system to reset itself if there is no operator response, and E is the total number of events that occurred in that trial. The method to determine the same type of percentage for the tracking and resource management tasks was different; however, the final answer is still in terms of the time the task was in an incorrect state versus the time the task could have been in an incorrect state. For the tracking task, the amount of time the cursor was outside of the box was compared to the total trial time, and for the resource management task, the amount of time the fuel level was outside of the specified range was compared to the total trial time. The final formula to determine the percentage of time the system was in the correct state is: $Y = 1 - average(X_S, X_R, X_C, X_T)$. Where: X_S is the incorrect percentage score for the system monitoring task, X_R is the incorrect percentage score for the resource management task, X_C is the incorrect percentage score for the communication task, and X_T is the

incorrect percentage score for the tracking task. An example of the Excel worksheets used to calculate these scores can be seen in Appendix C.

Data Analysis

To gain an understanding of how the heart-rate measures, the WWL, and the performance score are affected by the task load, the run number and the interaction of the two, a mixed model analysis of variance was employed. A mixed model is needed to account for the fact that the subjects are randomly chosen from an infinite population; therefore, they cannot be treated as fixed levels the way that the difficulty level and run number are. The linear statistical model is $y_{ijkl} = \mu + \tau_i + \beta_j + \omega_k + (\tau\beta)_{ij} + (\tau\omega)_{ik} + (\beta\omega)_{jk} + (\tau\beta\omega)_{ijk} + \epsilon\_ijkl$. Here τ_i is the random effect the subjects have on the experiment, β_j is the difficulty level, ω_k is the run number, the interaction $(\beta\omega)_{jk}$ is the interaction between difficulty levels and the run number, the interactions $(\tau\beta)_{ij}$, $(\tau\omega)_{ik}$, and $(\tau\beta\omega)_{ijk}$ are assumed to be random effects, and ϵ_{ijkl} is random error.

ANOVA Results

WWL and Performance Scores

The first measure investigated was the WWL, which is the weighted worked level calculated from the NASA TLX questionnaire. To determine if it was an accurate measure of the task load, six WWL scores were computed for each subject, one for each of the three workload levels, and each level was repeated, resulting in 78 samples. The ANOVA was performed with JMP software using the Residual Maximum Likelihood (REML) method, and any value p-value less than 0.05 was assumed significant.

These results show that the task load is the only significant fixed effect (F(3,25)=31.386, p<=0.0001). Performing a pairwise comparison with a Tukey's honestly significant difference (HSD) test shows that there is a significant difference between all task load levels except the two medium levels. Figure 9 shows workload level as a function of task load. As shown in this figure, the subjective estimate of workload (WWL) increased from 44.63 to 57.12 to 62.71 to 74.71 as the task load increased from the low through the high condition. The fact that significant differences in WWL exist only as a function of task load was expected because it is a more universal measure of workload. Because WWL wasn't significantly different between subjects or between runs, it shows that the measure was relatively stable and not affected as the subject gained experience with the software.

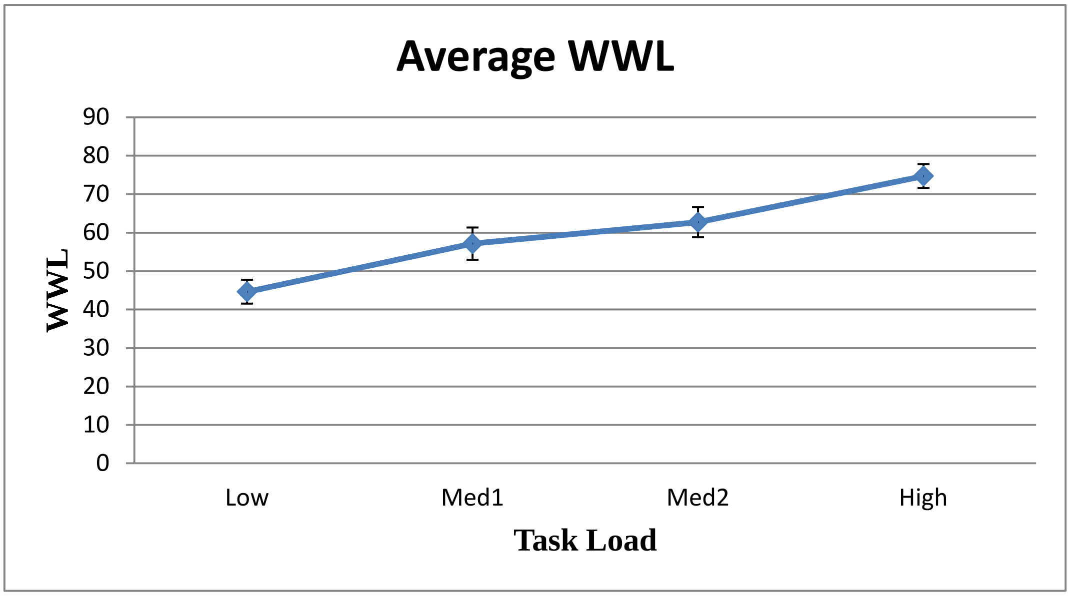

Figure 9: WWL by Task Load

The next measure investigated was the performance score. The JMP results show that both task load (F(3,24)=12.53, p<=0.0001) and run number (F(1,13)=12.14,

p<=0.0039) had a significant effect on performance scores. Performing a pairwise comparison with Tukey's HSD test shows that there was a significant difference in performance scores between low task load and the higher medium and high task loads, and that the high task load is significantly different than the lower medium and low task loads. There was no significant difference between the two medium levels of task load. Figure 10 shows performance score as a function of task load. As shown in this figure, the average performance score decreased from 0.65 to 0.57 to 0.55 to 0.48 as the task load increased from the low through the high condition.

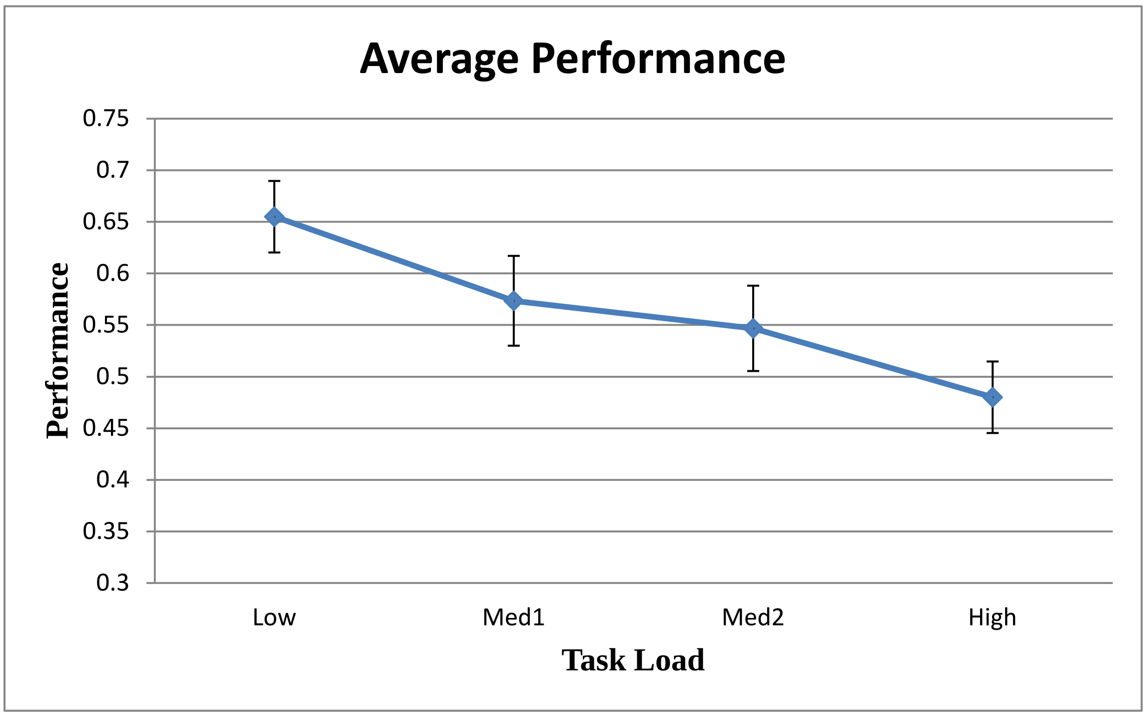

Figure 10: ANOVA on Performance Score

Figure 11 shows that the average performance scores also increased 5 percent between runs. This may indicate that the subjects would have benefitted from additional training. The performance score increased by 6.3 percent for this condition, indicating

that the subjects performance did not increase for these lower task loads as they gained experience between runs 1 and 2.

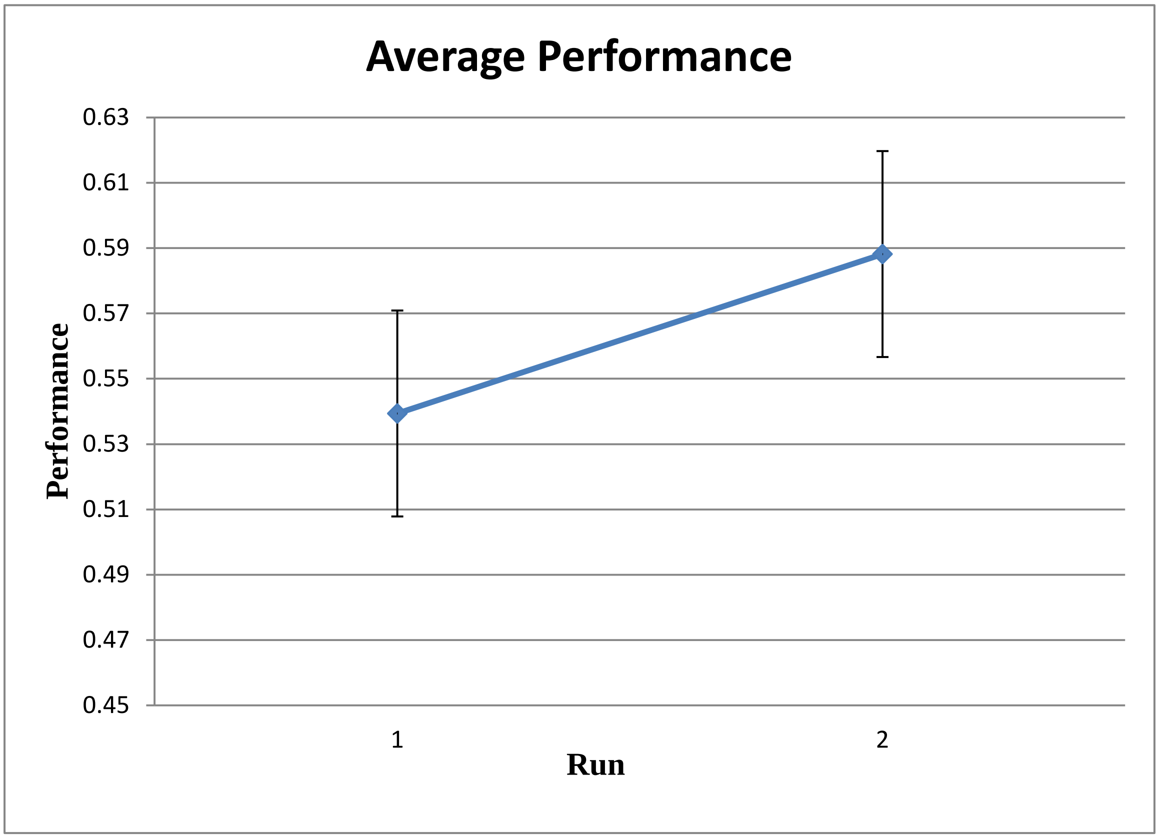

Figure 11: Pairwise Comparison of Task Load by Run

The random effect of subject produced much more variance for performance scores than it did for the WWL scores. This makes sense as some subjects will be more inclined to performing these types of tasks, while others may struggle greatly even with low task loads. These differences produce large variances in performance scores between subjects making it difficult to point to a specific performance score as being indicative of high mental workload for a subject.

Heart Rate Measures

ANOVA models for each of the heart rate measures were conducted with JMP using the previously stated method.

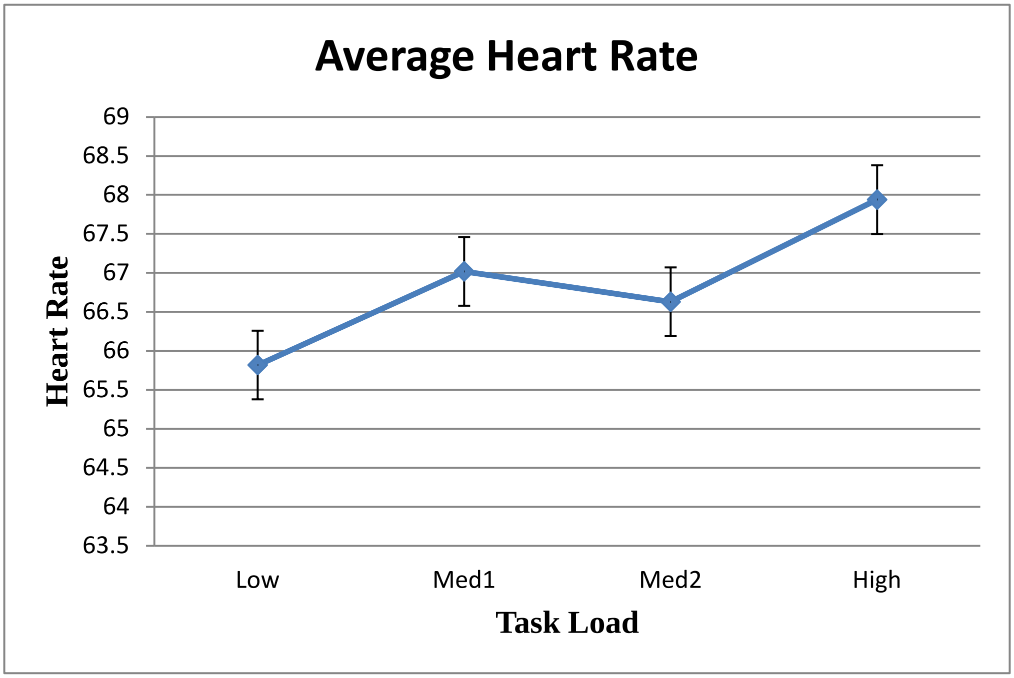

Figure 12: Heart Rate by Task Load

Heart rate was only significantly affected by task load (F(3,24)=6.34, p<=0.0026); however, it was only significant between the low and high task loads, which indicates that heart rate may only be able to discriminate between task loads with more extreme differences. The variance between subjects was also very high, which indicates that this is not a global measure but must be calibrated for each subject.

Performing the analysis on the low frequency (LF) component of heart rate variability (HRV) yielded significant effects for task load (F(3,23.43)=6.98, p<=0.0016) and run number (F(1,12.76)=5.2951, p<=0.0389). Figure 13 shows that the HRV (LF) value was highest at the low task load and lowest at the highest task load which was

expected; however, the two medium levels were neither significantly different from each other nor from the low or high levels. This may be due to the fact that the medium levels have half the number of data points and because subjects, who introduce a high level of variance to the ANOVA model, only completed one of the medium levels.

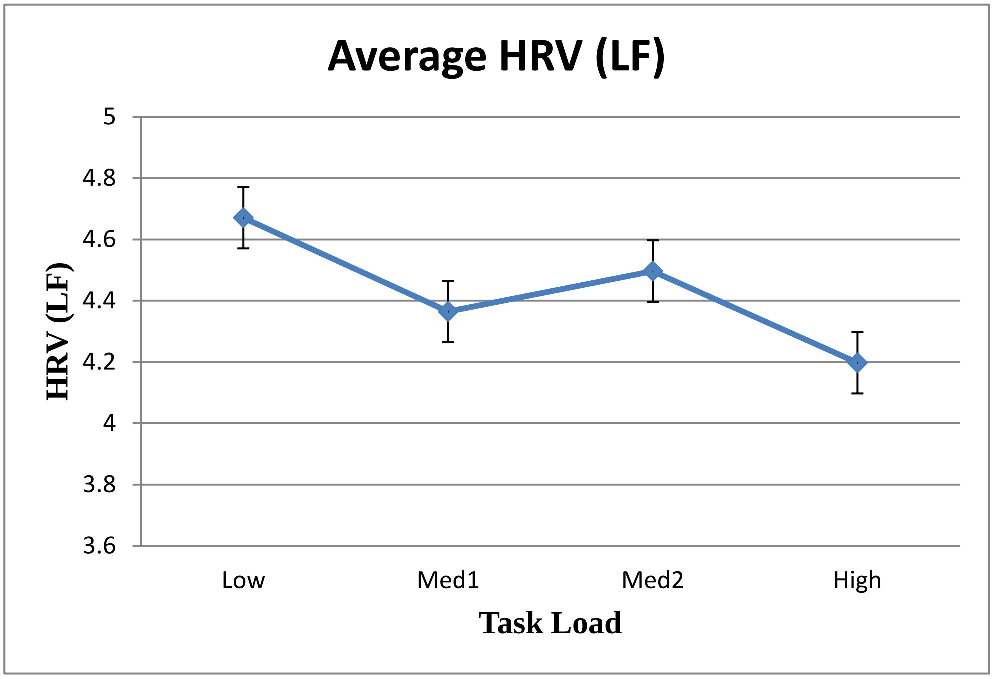

Figure 13: HRV (LF) by Task Load

The HF component of HRV, the SDNN, the HRV delta, and the HR delta were all significantly affected by task load. HRV (HF) was also significantly affected by run number and the interaction between run number and task load. Each of these measures showed trends similar to those seen for HR and HRV (LF). The CVRR measure did not show significant differences based on any of the fixed effects. See Table 1 for results.

Table 1: Heart Rate Measures with Significant Effects

HR Measure	Significant Effects	F Statistic	P-value
HF	Task Load Run Task Load x Run	F(3,22.6)=3.8279 F(1,15.77)=6.988 F(3,28.58)=3.3093	0.0235 0.0179 0.0341
SDNN	Task Load	F(3,23.39)=3.53	0.0305
CVRR	None	Not significant	Not significant
HRV Delta	Task Load	F(3,23.4)=3.72	0.0254
HR Delta	Task Load	F(3,23.7)=6.298	0.0027

Some interesting trends can be seen from all of the heart rate measures. The first is that all measures, except for CVRR, were significantly different for the low and high task loads, but were not significantly different with regards to the medium task loads. The responses for these task loads had significant variance. For some subjects they followed the expected trends (i.e. HRV decreased as task load increased and HR increased with task load); however, for some subjects the heart rate measures for medium task load did the opposite. A graph of each subject's LF for each task load can be seen in Appendix B, which highlights these differences. According to the ANOVA results, it appears that most of the heart rate measures can be used to determine when a subject is in a low workload situation or a high workload situation; however, they do not appear to be reliable for smaller changes. Luckily, performance scores for almost every subject show the expected trends; therefore, by using a combination of these measures it may be possible to more accurately determine a subject's workload without requiring them to complete any questionnaires.

Model building

To determine which measures would be most appropriate for determining a subject's perceived workload it was decided to use the Weighted Workload Level as the response since it represents the subject's perceived workload. Also, it was the only measure that accurately distinguished between each task load. The first step was to look at the correlations and partial correlations (the correlation with respect to all other correlations) between each measure and determine which attributes correlated most strongly with WWL. Figure 14 shows that performance scores are more strongly correlated with WWL than any of the heart rate measures. The partial correlations show that HR delta from baseline shows the next highest correlation.

Correlations

	Heart Rate	HRV (LF)	HRV (HF)	SDNN	CVRR	Performance score	WWL	HRV Delta	HR Delta
Heart Rate	1.0000	-0.8867	-0.7100	-0.6313	-0.5142	-0.1274	0.0471	-0.0649	0.3824
HRV (LF)	-0.8867	1.0000	0.7840	0.7608	0.6708	0.1265	-0.0972	0.2456	-0.4351
HRV (HF)	-0.7100	0.7840	1.0000	0.8314	0.7716	0.0317	0.0205	0.2131	-0.3913
SDNN	-0.6313	0.7608	0.8314	1.0000	0.9868	0.0099	0.0004	0.1706	-0.3664
CVRR	-0.5142	0.6708	0.7716	0.9868	1.0000	-0.0125	0.0032	0.1570	-0.3385
Performance score	-0.1274	0.1265	0.0317	0.0099	-0.0125	1.0000	-0.3289	-0.0399	-0.0231
WWL	0.0471	-0.0972	0.0205	0.0004	0.0032	-0.3289	1.0000	-0.1990	0.3268
HRV Delta	-0.0649	0.2456	0.2131	0.1706	0.1570	-0.0399	-0.1990	1.0000	-0.5841
HR Delta	0.3824	-0.4351	-0.3913	-0.3664	-0.3385	-0.0231	0.3268	-0.5841	1.0000

There are 1 missing values. The correlations are estimated by REML method.

Partial Corr

	Heart Rate	HRV (LF)	HRV (HF)	SDNN	CVRR	Performance score	WWL	HRV Delta	HR Delta
Heart Rate	.	-0.5160	0.1126	-0.6519	0.6806	0.0043	-0.0416	0.4846	0.3465
HRV (LF)	-0.5160	.	0.0549	0.1366	-0.0892	0.0757	-0.1135	0.3212	0.1020
HRV (HF)	0.1126	0.0549	.	0.4161	-0.3522	-0.0017	0.0934	0.0008	-0.1086
SDNN	-0.6519	0.1366	0.4161	.	0.9927	0.0201	0.0463	0.2451	0.2227
CVRR	0.6806	-0.0892	-0.3522	0.9927	.	-0.0295	-0.0430	-0.2685	-0.2380
Performance score	0.0043	0.0757	-0.0017	0.0201	-0.0295	.	-0.3388	-0.0687	0.0756
WWL	-0.0416	-0.1135	0.0934	0.0463	-0.0430	-0.3388	.	0.0016	0.2897
HRV Delta	0.4846	0.3212	0.0008	0.2451	-0.2685	-0.0687	0.0016	.	-0.5872
HR Delta	0.3465	0.1020	-0.1086	0.2227	-0.2380	0.0756	0.2897	-0.5872	.

partialed with respect to all other variables

Figure 14: Correlations and Partial Correlations

By building a linear regression model with WWL as the response and performance score and HR delta as the predictor variables the resulting R^2 value is 0.21 and the R^2 adjusted is 0.189, meaning that at most these two variables account for 18.9 percent of the variation in WWL. A stepwise regression function in JMP was applied to search for an improved model. The resulting model added HF as a predictor variable within the model, yielding an Adjusted R^2 value of 0.21. See Figure 14 for a regression plot showing the actual WWL values compared to the values of WWL that the model predicted.

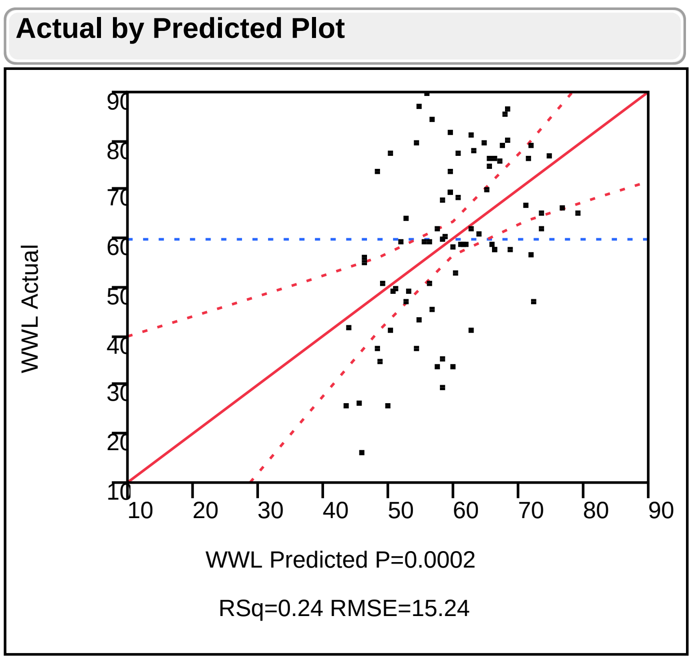

Figure 15: Actual WWL vs Predicted WWL

Because there was so much variation with the medium task load levels, the model was then examined with only the low and high task loads included in the model. The resulting correlations can be seen in Figure 16.

Correlations

	Heart Rate	HRV (LF)	HRV (HF)	SDNN	CVRR	Performance score	WWL	HRV Delta	HR Delta
Heart Rate	1.0000	-0.8794	-0.6962	-0.6042	-0.4768	-0.1031	0.1378	-0.0931	0.4686
HRV (LF)	-0.8794	1.0000	0.7735	0.7406	0.6420	0.1906	-0.1768	0.2309	-0.4558
HRV (HF)	-0.6962	0.7735	1.0000	0.8079	0.7450	0.0012	-0.0065	0.2114	-0.3947
SDNN	-0.6042	0.7406	0.8079	1.0000	0.9856	0.0234	-0.0649	0.1766	-0.3690
CVRR	-0.4768	0.6420	0.7450	0.9856	1.0000	0.0008	-0.0539	0.1632	-0.3253
Performance score	-0.1031	0.1906	0.0012	0.0234	0.0008	1.0000	-0.4208	0.0214	-0.0336
WWL	0.1378	-0.1768	-0.0065	-0.0649	-0.0539	-0.4208	1.0000	-0.2605	0.3700
HRV Delta	-0.0931	0.2309	0.2114	0.1766	0.1632	0.0214	-0.2605	1.0000	-0.5672
HR Delta	0.4686	-0.4558	-0.3947	-0.3690	-0.3253	-0.0336	0.3700	-0.5672	1.0000

Partial Corr

	Heart Rate	HRV (LF)	HRV (HF)	SDNN	CVRR	Performance score	WWL	HRV Delta	HR Delta
Heart Rate	.	-0.4706	0.0589	-0.6696	0.6966	0.1609	0.0698	0.4424	0.3829
HRV (LF)	-0.4706	.	0.1252	0.1554	-0.1113	0.2514	-0.0467	0.2604	0.1235
HRV (HF)	0.0589	0.1252	.	0.3035	-0.2425	-0.0760	0.1333	0.0463	-0.0599
SDNN	-0.6696	0.1554	0.3035	.	0.9930	0.0804	0.0977	0.2087	0.1875
CVRR	0.6966	-0.1113	-0.2425	0.9930	.	-0.0934	-0.1036	-0.2266	-0.2007
Performance score	0.1609	0.2514	-0.0760	0.0804	-0.0934	.	-0.4164	-0.0916	0.0885
WWL	0.0698	-0.0467	0.1333	0.0977	-0.1036	-0.4164	.	-0.0885	0.2673
HRV Delta	0.4424	0.2604	0.0463	0.2087	-0.2266	-0.0916	-0.0885	.	-0.5727
HR Delta	0.3829	0.1235	-0.0599	0.1875	-0.2007	0.0885	0.2673	-0.5727	.

partialed with respect to all other variables

Figure 16: Correlations and Partial Correlations

The correlation values without the medium task load levels appeared to be greater; therefore, stepwise regression was once again implemented in JMP. The best model (See Figure 17), which included performance score and HR delta as variables, yielded slightly improved Adjusted R^2 value of 0.275.

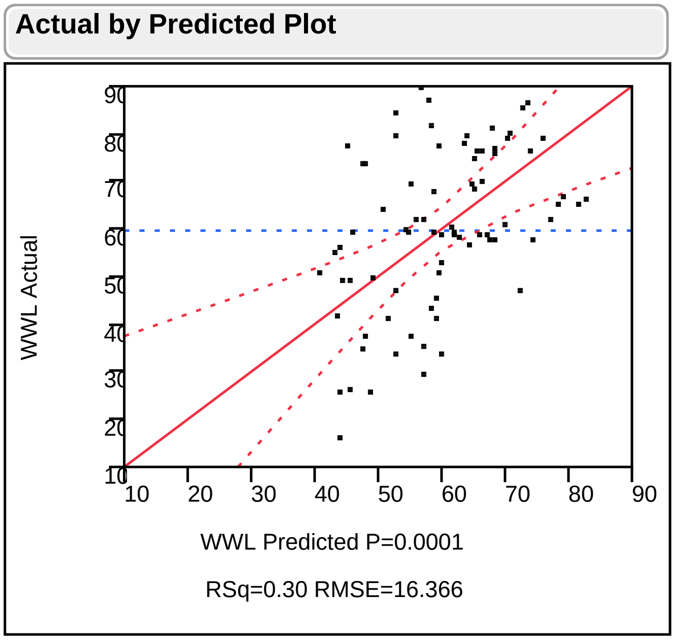

Figure 17: Actual WWL vs Predicted WWL

Because the HR delta was chosen for each model the data was also investigated with only the subjects who had a more controlled baseline reading (subjects 8-13). The stepwise regression of the model included WWL as the response variable and performance score, HR delta, HRV delta, and HF as predictor variables. The resulting model (See Figure 18) had an R^2 value of 0.61 and an Adjusted R^2 value of 0.55, which is a significant increase compared to the previous model, indicating the importance of a controlled baseline.

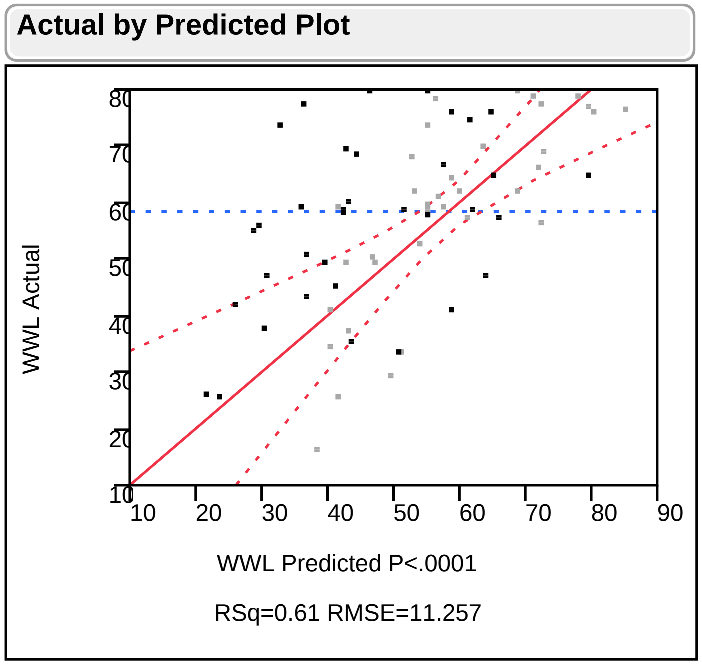

Figure 18: Actual WWL vs Predicted WWL

Discussion

Data were analyzed for multiple heart rate measures, performance scores, and subjective workload scores using ANOVA methods. It was determined that all but one of the measures was able to distinguish between low and high task loads. Regression models were also built to predict subjective workload levels, and by using the data from the subjects with more accurately controlled baseline data, an Adjusted R^2 value of 0.55 was achieved. The findings show promise in applying performance scores and various heart rate measures to trigger automation in systems.

While most of the results were consistent with the background literature a couple of notable differences stood out. In two studies CVRR was shown to be a significant indicator of mental workload (Verwey & Veltman, 1996; Kawakita, Itoh, & Oguri, 2010). However, the current study found no significant effects for this measure. An interesting note is that both of those studies were dealing with driving tasks, which may have a different effect on the various measures. Another result that slightly differed from other studies was that of HRV (LF). While this research did find that it was sensitive to high and low task load differences, it was not more sensitive than other measures as in other studies (de Waard, 1996).

One interesting finding for this research came from the HR delta and HRV delta results. Both of these measures showed results similar to the other heart rate measures when performing the ANOVAs. However, when investigating potential regression models that could be used to predict WWL, these measures had a more significant effect than the other heart rate measures. Because the regression model is looking to develop a model that can be used across subjects a delta measurement may provide a more accurate way to trigger adaptive automation.

V. Conclusions and Recommendations

Chapter Overview

Chapter 5 begins with a summary of the research along with some of the background. Each investigative question from Chapter 1 is then discussed based on results of the current research effort. While the first two research questions are answered in this research the question related to hyperspectral imaging was researched in

conjunction with Lt Elizabeth Norvell, an AFIT student in ENG, and her results will be referenced. Next, the significance of the research, especially in regard to adaptive automation, is discussed. Finally, recommendations for future research are given.

Research Overview

This research investigated the concept of using various performance and heart rate measures to trigger adaptive automation. Research has shown that adaptive automation can increase overall performance, and also that certain heart rate measures are indicative of an operator’s mental workload. A human-subjects experiment was developed in conjunction with a student from ENG that looked to gain more understanding of how various heart rate measures and performance measures could be used in adaptive automation. The results suggested that certain measures can be used to distinguish between task loads, and that a combination of measures can be used to model an operator’s perceived mental workload to some extent. However, the high variation between subjects makes it difficult to identify any one value within the various measures that might indicate when an operator is task saturated. While there are other physiological methods that have shown success as adaptive automation triggers, each of those methods are intrusive and not practical for day-to-day operations. Several imaging techniques have been developed to remotely measure heart rate; however, each has its own set of limitations that makes it difficult to implement in many operations. Therefore, this research also sought to identify these limitations and determine some of the effects they would have in a real-world system implementation while identifying elements that would be required for successful implementation.

Answers to Investigative Questions

Question 1: What are the heart rate measures that indicate an operator has reached task saturation?

According to the results each of the heart rate measures investigated, excluding CVRR, can be used to distinguish between large task load differences. However, none of the heart rate measures were able to accurately distinguish between the medium levels of task load. This could be an issue for successful implementation of adaptive automation, because these measures may only be able to detect the change after the operator has become task saturated causing decreased performance. Another issue that arises with each of these heart rate measures is the significant variance introduced by the subjects, making it difficult, if not impossible, to identify a specific value or change in heart rate that can be used globally to determine an operator's task load and trigger the appropriate automation.

While it may not be possible to use a specific heart rate measure across subjects, there are signs that a model can be used to estimate an operator's perceived mental workload is for a task. The performance score that was developed was able to provide additional distinction between task load levels; therefore, by using this score and a combination of the various heart rate measures a regression model was developed for this research that accounted for: 21% of the variance for all subjects and all task loads, 27.5% of the variance for all subjects and only the low and high task loads, and 55% of the variance for the 6 subjects with a controlled baseline and all task loads. These models would allow an adaptive automation system to estimate an operator's workload, which can then be refined according to an individual's performance and heart rate.

Question 2: What are the variables affecting the robustness of heart rate measurement techniques and what are some of the requirements to implement these techniques in an adaptive system?

A limiting factor in implementing the use of physiological measures to determine an operator's mental workload in normal operating conditions is the intrusiveness of these techniques. Therefore, the development of a robust technique to remotely measure an operator's heart rate is desirable. A number of remote imaging techniques have been developed that can accurately determine an individual's heart rate; however, these techniques often have limitations that may not allow them to be implemented in an actual system. Some of these limitations are related to specific subjects (e.g. large head movements or skin tone) while others are specific to the types of equipment employed (e.g. data rates and the need for dedicated light sources). Also, some of these limitations may affect simply obtaining a signal containing an individual's heartbeat, while other factors only affect HRV measures. Therefore, it is important to identify the variables affecting the technique's robustness and to determine what the requirements are for successful system implementation.

Variables Affecting Robustness

The variables that affect an imaging technique's robustness can be divided into three main categories: sensor, operator, and environment. Because imaging sensors are looking for changes in reflectance intensity there are several factors that can influence their usefulness. Some of the factors that affect sensors are: the wavelengths of light that are examined, the distance they are from the operator, interference from other light sources, operator movement, and the viewing angle. Other variables include the data rate

at which sensors can sample images and the manner in which the images are processed to obtain a signal identifying the heartbeat.

As with the human subjects testing for this research, the operator in a system introduces a lot of variables that can affect the performance of remote imager measurements. Some variables affecting remote imaging that are inherent to an individual are: skin tone, sweat, facial hair, and skin blemishes, such as birth marks. How much an operator moves their face can also affect the signal received by the imager. Different adornments, such as glasses and tattoos, are also variables to consider when determining an imager's requirements for implementation.

Different environmental factors also affect a technique's robustness. Temperature can affect the observed signal, as can humidity levels. Ambient lighting features can also either positively or negatively affect the reflectance signal received by the imager. Also, there is a big difference in implementation if an operator is at one console as compared to an operator moving throughout a room or an operator being monitored in a vehicle.

Requirements

There are several requirements that can be determined both by literature reviewed for this research and from the actual research. One requirement for maintaining a consistent signal is ensuring that signals are obtained from the same area of interest on an operator for the duration of the recording. While this is obviously not possible in an operational setting with only one imager, it may be possible to implement an array of synchronized imagers that employ effective face tracking. This would allow the same area of interest on an operator to be imaged in an operational system over an extended period of time. This measure is important for obtaining an accurate heart rate and

becomes even more important when determining any of the HRV measures. For example, a fairly accurate heart rate can be obtained after only 10 seconds and becomes very accurate with a minute of recording; whereas to obtain accurate measures of HRV every signal over a 2-minute period must be captured for the spectral domain and 5 minutes of accurate recording is required for time domain measures.

The accuracy of recording the beats also becomes more important for HRV as these measures are derived from the inter-beat intervals (IBI) which are generally measured in milliseconds. This places a requirement on the data rates of the imagers being used in the system. Current ECG systems can record heart data at very high sampling rates (1000 samples per second for this research). While 1000 Hz is not necessary for obtaining accurate results for some HRV measures such as SDNN most literature recommends at least 250 Hz to obtain a clean ECG waveform. However, most imaging systems cannot capture data at this rate. The need for accuracy can be seen by looking at the SDNN data obtained for this research. When analyzing that measure for significance differences, standard deviations of only 8 milliseconds were found to be significant. However, when looking within subjects many of the differences between task load levels is observable at 10 millisecond differences. This would make sample rates of at least 100 Hz feasible for measuring at least some HRV components though some accuracy would definitely be lost.

Another requirement for the sensor is the ability to accurately detect heartbeat signals for operators with various skin tones. Studies have shown that different reflectance curves are returned for various skin tones (Wueller, 2009); therefore, it

becomes important to determine whether or not the imaging technique can correctly interpret the various signals.

The previous requirements focused on the actual sensor, but there will also need for requirements to be placed on the operational environment to ensure success. Many imaging techniques require a dedicated light source, which would not be feasible in operational settings; however, some success with imagers has been found with ambient lighting. As signal detection is improved with these ambient lighting techniques it may become possible to accurately detect heartbeat, but there will still need to be lighting requirements for the environment. One of these would include the requirement for consistent luminance levels throughout the room with adequate lighting from every direction, which would minimize shadows on an operator's face.

Question 3: Can we develop a robust hyperspectral technique to remotely detect heart rate?

This question was considered to try to negate some of the shortcomings from the other techniques. The idea being that hyperspectral imaging techniques may be implemented to identify more discriminating features, which can aid in heartbeat detection in a larger range of scenarios. This research was performed in conjunction with Lt Elizabeth Norvell, whose thesis provides more complete information on this topic.

Significance of Research

This research was able to determine a variety of heart rate measures that can distinguish between task load levels, which can aid in developing triggers for adaptive automation. A global performance score for the AF_MATB was developed which will

allow for easier performance comparison between subjects and between task load levels. This score was also shown to be a better discriminator than physiological measures in terms of a subject's perceived mental workload; therefore, it can be used to further study adaptive automation in the AF_MATB. Implementing the results from these findings can allow for more effective adaptive automation that will allow for increased mission performance.

Another significant part of this research was identifying the variables that can affect imaging techniques. By identifying these variables it was possible to determine requirements for successful implementation into adaptive systems. This provides further information for researchers on whether or not a certain technique can be implemented, or if significant changes will be needed. Considering these requirements early on may aid in decisions early in a project's development.

Recommendations for Future Research

While this research was able to provide more insight into the usefulness of various heart rate measures and identify requirements for implementation of remote imaging techniques, there are still areas to be investigated for successful implementation into adaptive systems.

One recommendation for future research would be to conduct an experiment in which adaptive automation is triggered by a combination of the performance scores and heart rate measures that were chosen for the regression model. This would allow researchers to not only determine the effectiveness of the adaptive automation (e.g. by

comparing the performance scores), but would also allow for a better understanding of the model's accuracy while identifying ways to improve it.

There may also be benefit in looking at models besides a linear regression model. A second or third order model may do a better job at predicting an operator's mental workload.

I would also recommend only using one moderate level of task load and broadening the range of the task loads by decreasing the difficulty level. For the current research I do not believe the low task load setting was low enough for the subjects to experience underload.

Another recommendation would be to provide more thorough training prior to the study so that performance scores do not fluctuate between runs based only on experience with the software. This would also give a better understanding of whether or not heart rate measures act similarly after a learning plateau is reached. A way to incentivize subjects would also be beneficial as it helps to better mimic real-world scenarios where an operator's job may depend on their performance. A subject's motivation can affect both their performance scores and their heart rate measures.

Finally, there is a lot of room for further research into methods and techniques to remotely image heart rate, while looking for solutions to some of their shortcomings. While it may remain impossible to record heartbeats accurately enough to determine HRV remotely, there are definitely possibilities for using contact PPG methods that could be integrated into helmets in aircraft or ground vehicles that are not any more intrusive than the equipment applied in existing operational environments. To this end, it may be

beneficial to determine the locations on an operator that provides the best signal while remaining relatively unobtrusive.

Summary

The current research was able to use human subject testing to identify key heart rate measures that can distinguish between task loads in subjects. Requirements for robust remote imaging techniques were also developed to aid future work and hyperspectral imaging techniques were investigated in partnership with ENG to aid in choosing and implementing remote imaging systems. Through integration of these findings more reliable and unobtrusive adaptive systems can be developed.

Bibliography

Averty, P., Athenes, S., Collet, C., & Dittmar, A. (2002). Evaluating a New Index of Mental Workload in Real ATC Situation Using Psychophysiological Measures. *IEEE* .

Berntson, G. G., & Stowell, J. R. (1998). ECG artifacts and heart period variability: Don't miss a beat! *Psychophysiology* , 127-132.

BIOPAC System, I. (2000). Application Notes. *Heart Rate Variability--Preparing Data for Analysis Using AcqKnowledge* . Goleta, CA, USA: BIOPAC Systems, Inc.

Booher, H. R. (2003). *Handbook of Human Systems Integration.* Hoboken, NJ: John Wiley and Sons, Inc.

Corral, L. F., Paez, G., & Strojnik, M. (2012, June 18). Optimal wavelength selection for non-contact reflection photoplethysmography. *SPIE* .

de Waard, D. (1996). *The Measurement of Drivers' Mental Workload.* Haren, The Netherlands: The Traffic Research Centre VSC, University of Groningen.

Fedorovskaya, E. A., Miller, P., Prabhu, G., Horwitx, C., Matraszek, M., Parks, P., et al. (2001). Affective imaging: Psychological and physiological reactions to individually chosen images. *SPIE 4299* , 524-532.

Kamshilin, A. A., Miridonov, S., Teplov, V., Saarenheimo, R., & Nippolainen, E. (2011, April 1). Photoplethysmographic imaging of high spatial resolution. *BIOMEDICAL OPTICS EXPRESS* .

Kawakita, E., Itoh, M., & Oguri, K. (2010). Estimation of Driver's Mental Workload Using Visual Information and Heart Rate Variability. *13th International IEEE*, (pp. 765-769). Portugal.

Miller, W. D. (2010). *The U.S. Air Force-Developed Adaptation of the Multi-Attribute Task Battery for the Assessment of Human Operator Workload and Strategic Behavior.* WPAFB: AFRL.

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., & Yamamura, T. (2009). Physiological responses to workload change. A test/retest examination. *Applied Ergonomics* , 987-996.

Murai, K., & Hayashi, Y. (2011). Evaluation of a Navigator's Skill Based on Physiological Index. *2011 Fourth International Conference on Emerging Trends in Engineering & Technology* (pp. 135-139). IEEE.

Novak, D., Mihelj, M., & Munih, M. (2010). Psychophysiological responses to different levels of cognitive and physical workload in haptic interaction. *Robotica* , 1-8.

Parasuraman, R. (2003). Adaptive Automation Matched to Human Mental Workload. In G. R. Hockey, *Operator Functional States* (pp. 177-193). IOS Press.

Parasuraman, R., & Byrne, E. (2003). Automation and Human Performance in Aviation. In P. S. Tsang, & M. A. Vidulich, *Principles and Practice of Aviation Psychology* (pp. 311-348). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Parasuraman, R., & Hancock, P. A. (2008). Mitigating the Adverse Effects of Workload, Stress, and Fatigue with Adaptive Automation. In P. A. Hancock, & J. L. Szalma, *Performance Under Stress* (pp. 45-54). Burlington, VT: Ashgate.

Parasuraman, R., Cosenzo, K. A., & De Visser, E. (2009). Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness, and Mental Workload. *Military Psychology* , 270-297.

Phillips, C. A., Repperger, D. W., Kinsler, R., Bharwani, G., & Kender, D. (2007). A quantitative model of the human–machine interaction and multi-task performance: A strategy function and the unity model paradigm. *Computer in Biology and Medicine* , 1259-1271.

Reid, G. B., & Colle, H. A. (1988). Critical SWAT Values for Predicting Operator Overload. *Proceeding of the Human Factors Society*, (pp. 1414-1418).

Stanton, N. A., Salmon, P. M., Walker, G. H., Baber, C., & Jenkins, D. P. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design.* Burlington, VT: Ashgate Publishing Company.

Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. (1996). Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *American Heart Association, Inc* , 1-57.

Tattersall, A. J., & Fairclough, S. H. (2003). Adaptive Automation and Modes of Control. In *Operator Functional State* (pp. 238-246). Liverpool John Moores University, UK: IOS Press.

Tsang, P. S., & Vidulich, M. A. (2003). *Principles and Practice of Aviation Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates.

USAF. (2009). *United states air force unmanned aircraft systems flight plan 2009-2047*. Washington DC: Headquarters: United States Air Force.

Veltman, J. A., & Gaillard, A. W. (1996). *Physiological workload reactions to increasing levels of task difficulty*. TNO Human Factors Research Institute.

Verkruysse, W., Svaasand, L. O., & Nelson, J. S. (2008, December 22). Remote plethysmographic imaging using ambient light. *Optics Epress* .

Verwey, W. B., & Veltman, H. A. (1996). Detecting Short Period of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. *Journal of Experimental Psychology: Applied* , 270-284.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors* , 449-455.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering Psychology and Human Performance* . Upper Saddle River, NJ: Prentice-Hall Inc.

Wueller, D. (2009). In Situ Measured Spectral Radiation of Natural Objects. *17th Color Imaging Conference* (pp. 159-163). Frechen, Germany: Society for Imaging Science and Technology.

Appendix A

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

Appendix B

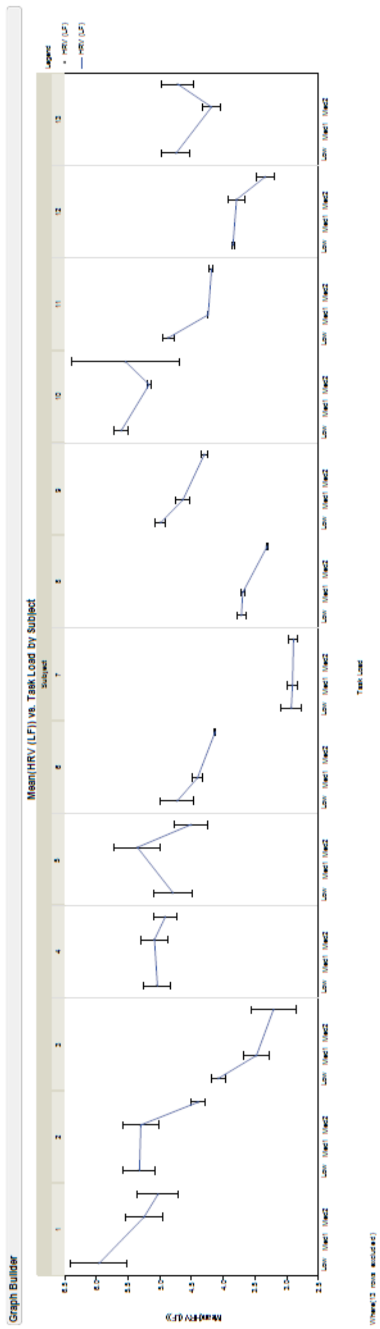

Figure 19: Subject's LF by Task Load

Appendix C

16										
17	Event Tim	X Coordin	Y Coordinates	X-Axis Difference	Y-Axis Difference	Time between events	X-axis out	Y-axis out	Out of box	Time out of box
18	0.158789	601.7771	568.6	3.777134	2.6	0.158789	0	0	0	7.480774
19	0.16809	605.4711	570.95664	7.471068	4.95664	0.009301	0	0	0	
20	0.198443	609.0828	573.31328	11.082842	7.31328	0.030353	0	0	0	
21	0.312449	612.6946	574.69544	14.694616	8.69544	0.114006	0	0	0	
22	0.397911	616.3064	575.67096	18.306391	9.67096	0.085462	0	0	0	
23	0.51224	619.9182	576.64648	21.918165	10.64648	0.114329	0	0	0	
24	0.60123	623.5299	577.622	25.529939	11.622	0.08899	0	0	0	
25	0.708356	627.1417	577.94752	29.141713	11.94752	0.107126	0	0	0	
26	0.797946	630.7535	578.02968	32.753487	12.02968	0.08959	0	0	0	
27	0.912275	633.7974	578.11184	35.797421	12.11184	0.114329	0	0	0	
28	1.001007	636.1914	578.194	38.191355	12.194	0.088732	0	0	0	
29	1.102518	638.5853	578.27616	40.585289	12.27616	0.101511	0	0	0	
30	1.201454	640.9792	578.35832	42.979224	12.35832	0.098936	0	0	0	
31	1.323833	643.5354	578.44048	45.535398	12.44048	0.122379	0	0	0	
32	1.399552	646.2538	578.52264	48.253812	12.52264	0.075719	0	0	0	
33	1.501232	649.6222	578.6048	51.622226	12.6048	0.10168	0	0	0	
34	1.608418	652.9906	578.68696	54.99064	12.68696	0.107186	0	0	0	
35	1.697699	656.3591	578.76912	58.359054	12.76912	0.089281	0	0	0	
36	1.808603	659.9708	578.85128	61.970828	12.85128	0.110904	0	0	0	
37	1.9022	663.5826	578.93344	65.582602	12.93344	0.093597	0	0	0	
38	2.019454	666.1388	579.0156	68.138777	13.0156	0.117254	0	0	0	
39	2.097761	667.2327	579.09776	69.232711	13.09776	0.078307	0	0	0	
40	2.207821	665.7266	579.17992	67.726645	13.17992	0.11006	0	0	0	

Subject 8_2_Tracking_Coordinate

Ready

Figure 20: Example of Excel Worksheet for Tracking Task Calculation

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From – To)*
21 March 2013	Master's Thesis	Aug 2011 –March 2013

4. TITLE AND SUBTITLE
Applying Hyperspectral Imaging To Heart Rate Estimation For Adaptive Automation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Splawn, Joshua M., Captain, USAF

5d. PROJECT NUMBER
13V251

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way, Building 640
WPAFB OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENV-13-M-30

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
711th Human Performance Wing RHCPA
Justin R. Estepp
2510 Fifth Street, Bldg 840
Wright-Patterson AFB, OH 45433-7022
798-3602 Justin.Estepp@wpafb.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)
711th HPW/RHCP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Automation use continues to increase in Air Force systems with the goal of improving operator efficiency and effectiveness. Unfortunately, systems are often complex, potentially imposing increased mental task load on the operator, or placing the operator in a supervisory role where they can become dependent on automation. Adaptive automation is a proposed solution, where automation is triggered when an operator is overloaded, and disabled as the operator is underloaded. Changes in physiological measures have shown promise in triggering automation. Unfortunately heart rate measurement can be obtrusive and impractical in day-to-day operations. This research used the Air Force Multi-Attribute Task Battery to impose varying task loads on subjects while monitoring their performance, recording their heart rate information with an electrocardiogram and obtaining subjective estimates of mental workload. Simultaneously, hyperspectral images were captured to determine if changes in heart rate might be identified through these images, providing a remote assessment of heart rate (HR). HR and several heart rate variability measurements where significantly affected by Task Load. A linear regression model was developed to predict subjects' perceived workload as a function of a proposed summary performance metric and HR measures. Additionally, this research identified several requirements for remote HR monitoring techniques.

15. SUBJECT TERMS
Adaptive Automation; Mental Workload; Hyperspectral Imaging; HRV

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Miller, Michael E., Ph.D
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	U	74	(937) 255-6565, x 4651 (Michael.miller@afit.edu)

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18